Journal of Hydrology 526 (2015) 151–163

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Probabilistic assessment of agricultural droughts using graphical models

Meenu Ramadas *, Rao S. Govindaraju

Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:
Available online 22 September 2014

Keywords:
Agricultural droughts
Probabilistic drought index
Crop water stress
Hidden Markov models

SUMMARY

Agricultural droughts are often characterized by soil moisture in the root zone of the soil, but crop needs are rarely factored into the analysis. Since water needs vary with crops, agricultural drought incidences in a region can be characterized better if crop responses to soil water deficits are also accounted for in the drought index. This study investigates agricultural droughts driven by plant stress due to soil moisture deficits using crop stress functions available in the literature. Crop water stress is assumed to begin at the soil moisture level corresponding to incipient stomatal closure, and reaches its maximum at the crop's wilting point. Using available location-specific crop acreage data, a weighted crop water stress function is computed. A new probabilistic agricultural drought index is then developed within a hidden Markov model (HMM) framework that provides model uncertainty in drought classification and accounts for time dependence between drought states. The proposed index allows probabilistic classification of the drought states and takes due cognizance of the stress experienced by the crop due to soil moisture deficit. The capabilities of HMM model formulations for assessing agricultural droughts are compared to those of current drought indices such as standardized precipitation evapotranspiration index (SPEI) and self-calibrating Palmer drought severity index (SC-PDSI). The HMM model identified critical drought events and several drought occurrences that are not detected by either SPEI or SC-PDSI, and shows promise as a tool for agricultural drought studies.

1. Introduction

The onset of an agricultural drought event is typically marked by a decline in the soil moisture level below a threshold value that affects crops. Precipitation, soil moisture, and temperature are the common variables adopted for agricultural drought studies (Mishra and Singh, 2010). Various indices for characterizing agricultural droughts are listed in Maity et al. (2013). Among these, Palmer Drought Severity Index (PDSI) (Palmer, 1965), Crop Moisture Index (CMI) (Palmer, 1968), Soil Moisture Anomaly Index (Bergman et al., 1988), and Vegetation Condition Index (VCI) (Liu and Kogan, 1996) are popular.

Meteorologic and hydrologic drought indices (e.g., Standardized Precipitation Index SPI, and PDSI) have been often used in agricultural drought studies (Narasimhan and Srinivasan, 2005). The PDSI uses both precipitation and surface air temperature as inputs, in contrast to SPI that uses precipitation alone. However, PDSI is limited as an indicator of soil moisture status or as being capable of identifying agricultural droughts; it demonstrates good correlation with soil moisture content during warm seasons but weak correlation in spring as the underlying model does not account for the effect of snowmelt (Dai et al., 2004). Palmer (1968) developed the Crop Moisture Index (CMI) as an index for short-term agricultural droughts from procedures similar to the PDSI. The CMI is computed from evapotranspiration deficits for monitoring short-term agricultural drought conditions that modulate crop growth. Meyer et al. (1993) developed a Crop Specific Drought Index (CSDI) for corn using evapotranspiration estimates. An alternative drought index—Standardized Precipitation Evapotranspiration Index (SPEI) that possesses the merits of PDSI and SPI in terms of sensitivity to temperature-driven evaporation that is important in crop growth and multi-scalar properties, respectively, was proposed by Vicente-Serrano et al. (2010). The performance of SPEI in drought impact analyses and climate change studies is well documented (Yu et al., 2013; Potop et al., 2012; Vicente-Serrano et al., 2010).

Researchers typically regard soil moisture as the most appropriate indicator of agricultural droughts (Keyantash and Dracup, 2002; Karamouz et al., 2004; Sheffield and Wood, 2008). Estimation of soil moisture from ground measurements is difficult due to heterogeneity caused by the spatially varying precipitation, land cover, soil and topography (Margulis et al., 2002; Vereecken et al., 2008). Temporal and spatial resolution of soil moisture is also crucial for predicting adequate soil profile wetting and drying

* Corresponding author.
E-mail addresses: mramadas@purdue.edu (M. Ramadas), govind@purdue.edu (R.S. Govindaraju).

http://dx.doi.org/10.1016/j.jhydrol.2014.09.026
0022-1694/

between precipitation events. The role of soil moisture in recurring droughts in North America was studied by Oglesby and Erickson (1989). Sheffield et al. (2004) used soil moisture estimates from the Variable Infiltration Capacity (VIC) model to develop a drought index that identified major drought events of the past and had good correlations with PDSI. Lakshmi et al. (2004) found that the deep layer soil moisture was capable of characterizing droughts in the Mississippi River Basin. The Soil Moisture Deficit Index (SMDI) developed by Narasimhan and Srinivasan (2005), based on weekly soil moisture deficits, had good correlation with indices such as SPI and PDSI, and offered better performance because of its fine spatial and temporal resolution. The authors used SWAT model to simulate daily soil moisture values at 4 km $\times$ 4 km spatial resolution that were then aggregated to a weekly time scale. Tang and Piechota (2009) explored the possibility of deep layer soil moisture as an indicator of climate extremes, and linked it to PDSI, precipitation, and streamflows. Their study utilized soil moisture as a drought indicator for characterizing the hydrologic status for the Colorado River Basin, and further identified the spatial and temporal variability of soil moisture in response to drought events in the region.

Root-zone soil moisture availability is used by agencies such as the United States Department of Agriculture (USDA)-International Production Assessment Division (IPAD)—as a major factor influencing crop yield forecasts (Bolten et al., 2010). When Wu et al. (2011) performed drought vulnerability assessment for China, seasonal crop water deficiency, available soil water-holding capacity and irrigation were adopted as the important drought indicators. The soil water holding capacity is a function of soil type, and varies spatially across a region creating patterns of crop water stress and water resource availability. Maity et al. (2013) characterized drought proneness of Malaprabha Basin, India, via a copula model for resilience and vulnerability values calculated from modeled soil moisture data for the region.

Since water needs vary with crops, agricultural drought incidences in a region can be assessed better if crop responses to soil water deficits are also accounted for in the index. Water stress influences rate of photosynthesis and stomatal closure, and affects crop production (Scholes and Walker, 1993). Denmead and Shaw (1960) studied the effect of soil moisture deficit on the development and yield of corn, by imposing soil moisture deficit at different growth stages. The changes in plant characteristics such as stalk height, cob length, area of the ear leaf, total production of stover and grain, and yield of grain under moisture stress were explored. Holt et al. (1964) investigated the effect of stored soil moisture at planting on corn yields, and developed regression equations for relating soil moisture to corn yield. A quantitative understanding of the plant response to water stress requires detailed study of soil moisture dynamics that include soil–water–air interaction, nutrient uptake by plants, and transpiration. Soil moisture deficits directly control the plant water potential that determines transpiration losses and the turgor pressure in plant cells (Porporato et al., 2001). The role of water stress in the structure and functioning of vegetation in African savannas (grassland ecosystems) was studied by Rodriguez-Iturbe et al. (1999a,b). The authors proposed a measure of "static" vegetation stress that can be calculated from soil moisture levels corresponding to plant wilting and full turgor. The "static" stress is zero when soil moisture is above the level of incipient stomatal closure (full turgor) and reaches a maximum value of one when soil moisture is at the wilting point of a plant. These two stages are based on the effects of water stress on plant physiology (Hsiao, 1973). Porporato et al. (2001) later introduced "dynamic" water stress to address the mean intensity, duration and frequency of soil moisture deficits. Laio et al. (2001) developed a stochastic model for soil moisture and water balance studies.

Drought conditions for crops in the Midwest are, by and large, determined by the soil water availability rather than by precipitation or evaporation. The plant response to water stress in the root zone of a soil could be used to develop a new agricultural drought index. Such an index would take due cognizance of crop needs. However, the changing soil moisture status and different crop rotation patterns followed in agricultural fields require that the drought analysis be performed in a statistical sense. A probabilistic assessment would convey the uncertainty in agricultural drought classification that popular indices (SPEI, PDSI, SPI) do not provide. Madadgar and Moradkhani (2013, 2014) developed a probabilistic forecast model for future hydrologic droughts in a Bayesian framework that allows probabilistic predictions and accounts for uncertainty in drought characterization. In this study, agricultural drought events in the state of Indiana are investigated in a probabilistic framework using graphical models—specifically hidden Markov models (HMMs)—given the temporal dependence that exists between drought states. The crop stress function values derived from soil moisture data are used to define agricultural drought states (1-near normal, 2-moderate drought, 3-severe drought, and 4-extreme drought).

Hidden Markov models have been used for solving numerous practical problems in speech processing (Leggetter and Woodland, 1995), signal processing (Crouse et al., 1998), genomics (Yau et al., 2011), tunneling design (Leu and Adi, 2011), meteorological studies (Hocaoğlu et al., 2010) and air quality modeling (Zhang et al., 2012). Mallya et al. (2013a) utilized HMMs to model meteorologic and hydrologic droughts. Many of these applications used Gaussian emission distributions (Leggetter and Woodland, 1995; Burget et al., 2010; Mallya et al., 2013a). Alternatively, atmospheric ozone levels were modeled using Gamma hidden Markov models by Zhang et al. (2012), and Sun et al. (2013) used HMMs with log-normal, Gamma and generalized extreme value (GEV) distributions to predict particulate matter concentrations.

Unlike previous studies (Mallya et al., 2013a; Zhang et al., 2012), the crop water stress function used in this study is bounded between [0,1], and as a result, previously utilized emission distributions are not suitable. This paper describes a new class of HMMs with beta emission probability distributions. These new models were used for developing probabilistic classification models for agricultural droughts in Indiana. The merits of HMM-based probabilistic agricultural drought index over SPI, self-calibrating PDSI and SPEI were investigated. The organization of rest of the paper is as follows: Section 2 describes the study area and data used, Section 3 explains the methodology adopted in the development of the probabilistic index, followed by results and discussion in Section 4, and finally the conclusions derived from the study are presented in Section 5. In addition, Appendix A provides derivations of equations used in the methodology.

2. Study area and data used

To examine the applicability of the graphical model, the state of Indiana, USA is chosen as the study area. Indiana is nationally ranked for agricultural production, major cultivated crops being corn and soybean. For instance, Fig. 1 illustrates the cultivation pattern followed in a small patch of land in Lake County in northern Indiana during the period 2000–2012, where corn and soybean are predominant. Crop rotation, fallow land, and double cropping practices have been adopted in this area. Winter wheat, alfalfa and pasture grass were grown as minor crops in alternate years. Livestock and dairy farming thrive on agriculture over such farmlands in Indiana and other Midwest states.

Unfortunately, droughts are common in the Midwest, and hamper the prospects of large yields from these farms. Consequences of the recent 2012 drought in US can be found in Mallya et al. (2013b)

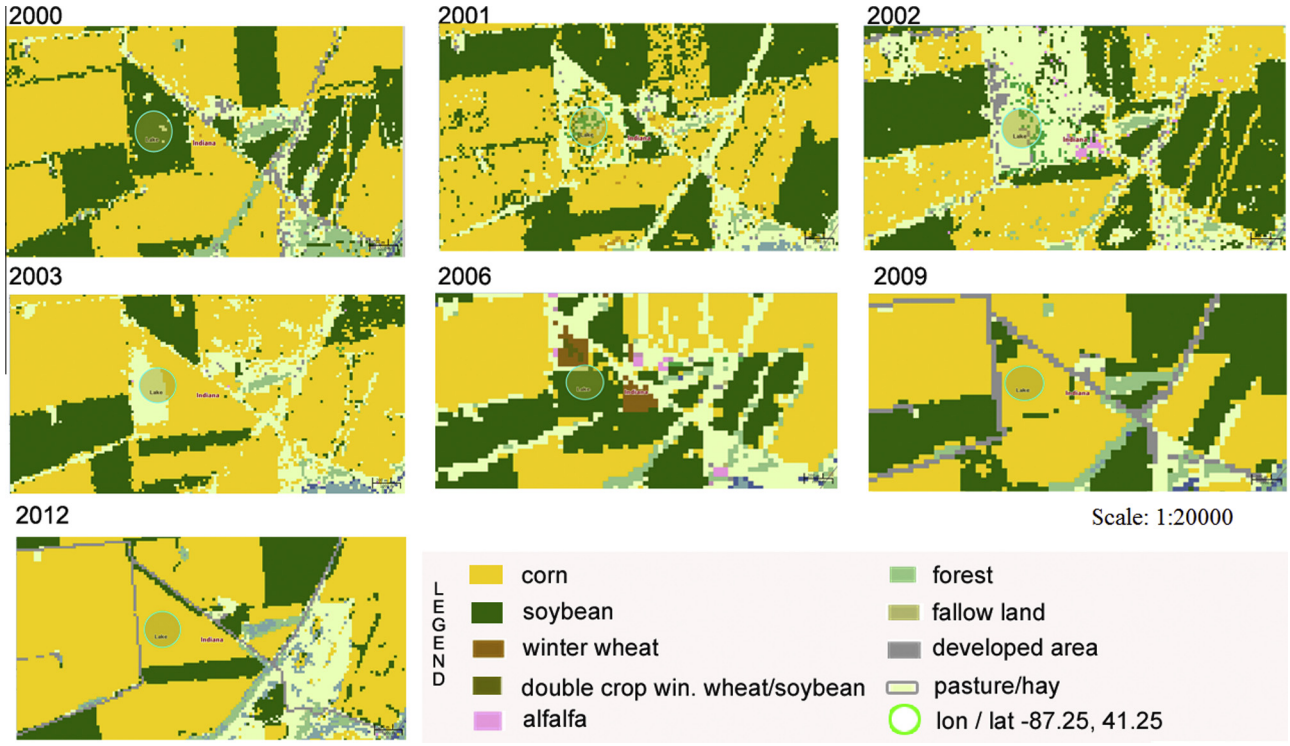

Fig. 1. Cropping pattern in a small patch of agricultural field in Lake County, Indiana, US during 2000–2012 where the yearly changes in land use and land cover are evident (adapted from http://nassgeodata.gmu.edu/CropScape/).

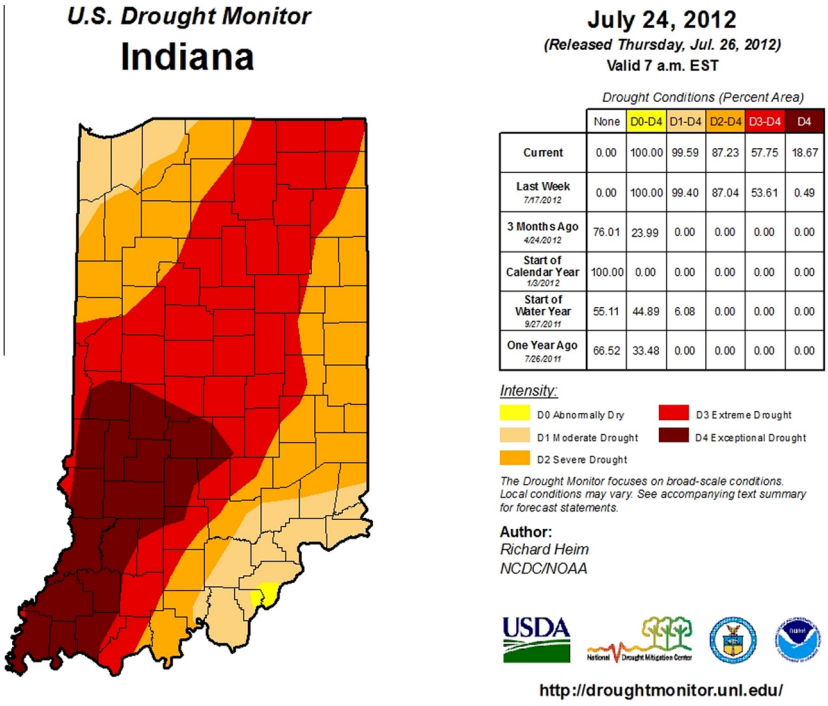

Fig. 2. Extent and magnitude of 2012 drought in Indiana- in July 2012, one of the driest months of the year, captured by the U S Drought Monitor. Droughts are labeled by their intensity: D0 being the least and D4 being the most intense. (The U.S. Drought Monitor is jointly produced by the National Drought Mitigation Center at the University of Nebraska-Lincoln, the United States Department of Agriculture, and the National Oceanic and Atmospheric Administration. Map courtesy of NDMC-UNL.)

and Kerr (2012). Fig. 2 shows the extent of drought extremes over Indiana evaluated by United States Drought Monitor (USDM) for July 24, 2012. The USDM map identifies regions experiencing different drought categories ranging from D0 (abnormally dry) to D4 (exceptionally dry) for that particular week, and the classification criteria are described in http://droughtmonitor.unl.edu/AboutUs/ClassificationScheme.aspx. More than half of the state was affected by an extreme drought (Fig. 2). The major impact of agricultural droughts is on crop cultivation in the affected regions. From an economic point of view, droughts have a detrimental effect on corn

and soybean prices in Indiana under the current agricultural conditions, and are particularly devastating to livestock producers (http://www.ibrc.indiana.edu/ibr/2012/outlook/articles/agriculture.pdf).

The yearly cropping pattern of Indiana was obtained from Cropland Data Layer (CDL) that is hosted on CropScape (Han et al., 2012; http://nassgeodata.gmu.edu/CropScape/). The CDL is a raster, geo-referenced, crop-specific land cover data layer created annually for the continental United States using moderate resolution satellite imagery and extensive agricultural ground truth. It is developed by the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA). This data is available from 2000 to 2012 for Indiana. Average crop distribution in acreage for this time window was extended to cover the 1948–2012 period.

For soil moisture data, the Climate Prediction Center's (CPC) $0.5^\circ \times 0.5^\circ$ resolution global monthly datasets (Fan and van den Dool, 2004) were used. The data sets have sufficiently long record lengths needed for robust modeling. Huang et al. (1996) outline the procedure for constructing monthly soil moisture time series data sets over the entire continental U.S. with a 1600 mm deep one-layer soil moisture model. Their model is based on the water budget in the soil and uses monthly temperature and monthly precipitation as inputs. Estimated evapotranspiration, runoff and groundwater loss used in the CPC soil moisture model are derived from these two inputs. A total of 52 CPC grid points fall over Indiana, and soil moisture data from the period 1948–2012 were extracted at these grid points for this study.

3. Methodology

Development of an HMM-based probabilistic drought index required estimation of crop water stress, studying the temporal dependence between drought states, choice of emission distribution, parameter estimation, and model selection. These various steps are briefly described in this section.

3.1. Estimation of crop moisture stress function

Plant water potential is controlled by the soil moisture present in the root zone. With excess moisture, the plant water potential increases and turgor in leaves is very high, as a result of which stomatal pores open and evapotranspiration is in full swing. However, under conditions of soil moisture deficit, there is a drop in the water potential in plants, inhibiting their ability to take up water from the soil, as a result of which the stomatal openings close to avoid loss of available water. The sequence of events that take place in plants in response to water stress can be understood based on the varying levels of stomatal closure. Incipient stomatal closure is among the first symptoms, and finally as the plant starts wilting, complete closure would take place. Rodriguez-Iturbe et al. (1999a,b) quantified the plant water stress as a function of soil moisture level in the soil at that instant ("static" water stress) ζ-such that it is zero when soil moisture is above the level of incipient stomatal closure and has its maximum value of 1 when the soil moisture deficit causes wilting (denoted as s^* and s_w respectively). Between s^* and s_w, the authors suggested a non-linear increase of plant water stress with soil moisture deficit as

$$\zeta(t) = \begin{cases} 1 & \text{for } s < s_w \\ \left[\frac{s^* - s(t)}{s^* - s_w}\right]^m & \text{for } s_w \leqslant s(t) \leqslant s^* \\ 0 & \text{for } s > s^* \end{cases} \quad (1)$$

where $s(t)$ is the soil moisture content at time t, and m is a measure of the non-linearity desired in the crop water stress model. A value of $m = 2$ is used for crops in this study. The values of s_w, s^* and m vary with plant species.

The crop distribution information at various CPC grid points over the Indiana region were extracted to develop corresponding weights for dominant agricultural crops and multiplied to the crop water stress value of each crop. The resulting monthly effective crop stress time series at each grid point was used for agricultural drought analysis in the region.

A hidden Markov model (HMM) was used to develop a probabilistic classification model to define agricultural droughts. A schematic of the graphical model used in the study is shown in Fig. 3. It illustrates the concept of estimating crop stress ζ using soil moisture and crop information. The non-linear increase in ζ between s^* and s_w is represented in the graph in Fig. 3. The HMM graph structure with the hidden drought states (in dashed boxes) is shown in the same figure. In this approach, a certain range of crop water stress values define a drought state, and the range varies spatially. Brief description of the theory of HMMs is provided in subsequent sections.

3.2. Temporal dependence in drought states

In the realm of statistical models, hidden Markov models are suitable for cases where temporal dependence in the drought states needs to be preserved. Otherwise, mixture models would suffice as a simpler tool for probabilistic modeling (Mallya et al., 2013a). Mutual Information (MI) statistic is used in this study to determine the nature of temporal dependence between drought states at one-month interval. The drought states are based on a standardized crop-drought index calculated using the crop stress function values.

Mallya et al. (2013a) provide a detailed analysis of the nature of temporal dependence between drought states for meteorological and hydrological droughts with durations greater than one month in which case use of a hidden Markov model was favored, and highlighted the merits of adopting a simpler Gaussian mixture model (GMM) when temporal dependence was insignificant. However, for soil moisture-driven droughts, Markovian dependence in time cannot be neglected without exploring the nature of dependence, as soil moisture holds a long-term persistent memory (Manabe and Delworth, 1990; Koster and Suarez, 2001). This aspect is investigated later in the paper.

3.3. Graphical models

A graphical model is a family of distributions that can be efficiently represented by a directed or undirected graph. Variables of interest are denoted by nodes whereas their dependencies are indicated by connections/edges. The graph structure allows users to compute marginal and joint conditional probabilities between variables present as nodes in the graph (Jordan, 2004). Graphical models have been popular in the fields of speech recognition, language processing, genetics, and information retrieval; recent applications include modeling spatial and temporal patterns of precipitation (Ihler et al., 2007), and extreme event modeling (Yu et al., 2012).

3.3.1. Hidden Markov models

Hidden Markov models are a class of graphical models where the graph structure comprises of hidden nodes with connections to observed nodes, such that temporal dependencies exist between the hidden nodes. In an HMM, as shown in Fig. 3, the outputs/observations of the system are assumed to be dependent on a sequence of hidden states. In the context of drought studies, the hidden nodes are the latent drought states, while the observations

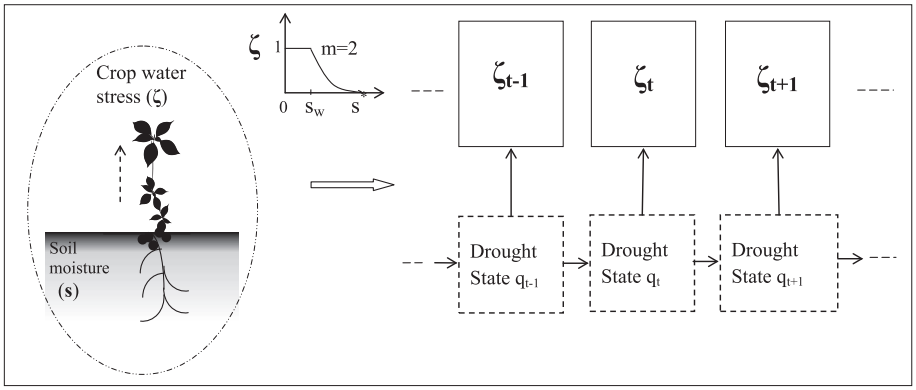

Fig. 3. A schematic of the HMM used in this study. The hydrologic variable ζ_t represents the crop water stress. The hidden drought state q_t represents one of near normal, moderate, severe or extreme drought states. The subscript t is the time index. ζ is estimated from soil moisture content values s, s_w (at wilting point) and s^* (at incipient stomatal closure), and m is the measure of non-linearity in the estimation of ζ_t.

may be precipitation or streamflow values (Mallya et al., 2013a), or soil moisture-driven crop stress values as in this study.

Consider the model where the hydro-climatic variable of interest at an instant t is denoted by x_t, $t = 1, 2, \ldots, N$ $\{x_t \in R$ and $X = [x_1, x_2, \ldots, x_N]^T = x_{1:N}\}$. The observation x_t is dependent on the hidden state variable q_t, $\{q = [q_1, q_2, \ldots, q_N]^T = q_{1:N}\}$ which is assumed to be a first order Markov process, i.e. the probability of the system being in any future state is independent of past states given the present state. The hidden variable q_t is considered a discrete variable representing one of the K possible latent states. The major characteristics of an HMM with K states and following first order Markov property are:

(i) Given the state of the system at time $t-1$, q_t is independent of previous states i.e. $P(q_t|q_{t-1}, q_{t-2}, \ldots, q_1) = P(q_t|q_{t-1})$. The state transition probability matrix can be defined as $A = \{a_{ij}\}$ where $a_{ij} = P(q_{t+1} = j|q_t = i)$, $1 \leqslant i, j \leqslant K$.
(ii) Given the current state q_t, the observation at that instant x_t is conditionally independent of past observations, and the probability $P(x_t|q_t)$ is known as the emission distribution. The matrix $B = \{\alpha, \beta\}$ represents the parameters of the emission distribution.
(iii) The initial state distribution, i.e., the probability of drought state at the instant $t = 1$ or $P(q_1)$ is given by $\pi = \{\pi_i\}$ such that $\pi_i = P(q_1 = i)$, $1 \leqslant i \leqslant K$.

Besides, the following constraints hold valid for the HMM.

$$\sum_{i=1}^{K} \pi_i = 1$$
$$\sum_{j=1}^{K} a_{ij} = 1; \quad 1 \leqslant i \leqslant K \tag{2}$$

The joint distribution of the different drought states and observations in the HMM can then be expressed as

$$P(q_{1:N}, x_{1:N}) = \pi_i \Pi_{t=2}^{N} P(q_t|q_{t-1}) \Pi_{t=1}^{N} P(x_t|q_t) \tag{3}$$

3.4. Model implementation

3.4.1. Emission distribution

Gaussian emission distributions have been favored in several continuous-HMM applications due to ease of computation. However, there are applications where Gaussian densities cannot be used, and hence, parameter estimation methods have to be designed from first principles. A beta emission distribution was adopted in this study for the following reasons: (i) it is a continuous distribution, (ii) it is well-suited for variates over the finite range of [0,1], (iii) it has the flexibility to model very skewed emission distributions that are needed for extreme events, and (iv) distributional parameters can be estimated in the HMM context.

3.4.2. Parameter estimation

An important task in generating an HMM-based probabilistic model for drought data is parameter estimation—finding the best set of $\{\pi, A, B\}$ such that the probability of the observation sequence given the model i.e., $P(O|\text{model})$ is maximized. Parameter estimation in HMM was performed using Baum-Welch algorithm that uses Expectation-Maximization (EM; Baum et al., 1970; Rabiner, 1989). The Baum Welch algorithm treats parameter estimation as a constrained optimization of $P(O|\text{model})$ subject to constraints in Eq. (2), and estimation formulae for $\{\pi, A, B\}$ are developed using a Lagrange multiplier technique such that the results yield maximum $P(O|\text{model})$ value. The details of parameter estimation including that for the shape parameters $\{\alpha, \beta\}$ of the emission distribution are provided in Appendix A.

The initial user-input values fed into the HMM framework play an important role in the estimation of probabilities and parameter values as the estimation algorithm may run into local maxima during the simulations. In order to ensure global optima are achieved, random sets of initial values were tried, and the estimated values corresponding to maximum probability $P(O|\text{model})$ were chosen for the model. Thus parameter estimation was a trial and error method. In scaled HMMs, the term $\log [P(O|\text{model})]$ is maximized (Rabiner, 1989).

Once the model parameters are estimated, the conditional probability of being in a particular drought state at time t, given the observations and set of model parameters is simply the posterior probability of falling in that state at time t (see Appendix A, Eq. (A11)). Probabilistic classification of drought states based on proposed crop water stress index is facilitated by estimating these probabilities using the HMM.

4. Results and discussion

4.1. Crop moisture stress estimation

Gridded soil moisture data at 52 locations over Indiana are used to compute the respective crop stress function values. Land cover data for these locations are retrieved from CDL provided by USDA-NASS. Only the major crops such as corn, soybean, sorghum, alfalfa, winter wheat, and double crops-winter wheat/soybean

(WS) and winter wheat/corn (WC) are considered in the drought analysis. The average acreage distribution of various crops grown in Indiana is as follows: 35–55% each of corn and soybean, less than 10% each of winter wheat and double crop WS, and less than 1% of sorghum, alfalfa and WC.

For all these crops, the water requirements over their growing seasons are assessed based on rooting depths at different growth stages (Evans et al., 1996). The adopted rooting depth variation with crop type and time of the year is shown in Table 1. Plant rooting depths were obtained mostly from past literature (Weaver and Bruner, 1927; Weaver, 1926; Rhoads and Yonts, 1991). Soil water content s_w at permanent wilting point (PWP) and s^* at incipient stomatal closure required for crop water stress calculation are computed as percentages of water available in the root zone of the crops, and these values are allowed to vary with different stages of plant growth. For instance, studies by Tolk (2003) determined PWP for corn and sorghum planted in 2-m deep soil to be around 488 mm and 420 mm respectively. For the different crops: soybean, alfalfa, and winter wheat, PWP, as a percentage of rooting depth are assumed to be 15, 10 and 19 percent respectively. The calculated monthly s_w and s^* values (in mm) for different crops are shown in Fig. 4. There is an increase in plant water requirement as the growth stage advances. These values are estimated based on the rooting depth values in Table 1 and plant water requirements mentioned previously. Under double cropping, values for s_w and s^* throughout the year are significant, unlike the case of a single crop as shown in Fig. 4. The crop stress function time series is computed for the growing season of crops. A weighted crop stress function time series is then calculated using crop acreage data at each grid location.

4.2. Exploring temporal dependence between drought states

Fig. 5 demonstrates the results of temporal dependence analysis conducted using mutual information statistic (MI; Cover and Thomas, 1991), where crop stress-based drought states for the month of January (as an example) are compared with those of other months. The crop stress function values are standardized and categorized similar to SPI-based drought classification (ranging from W4–W0, to normal to D0–D4; McKee et al., 1993). For instance, in a two bin case, W4-Normal and D0–D4 classes are grouped into two drought states: no-drought and drought respectively. In a similar fashion, the categories are grouped into 4 and 6 bins for estimating temporal dependence. For each of these cases, respective monthly MI statistics were computed using Eq. (4).

$$\mathrm{MI}(X,Y) = \sum_{x\in X}\sum_{y\in Y} p_{x,y}(x,y)\log\frac{p_{x,y}(x,y)}{p_x(x)p_y(y)} \tag{4}$$

where $p_{x,y}$, p_x, p_y are joint probability of (X, Y), and marginal probabilities of X and Y respectively. As an example, mutual information statistic values between drought states in January (X) and those in the rest of the months of the year (Y) were calculated from monthly time series of ζ for one station, and are plotted in Fig. 5. It is seen that the January drought states share temporal dependence with those of February and March, based on higher MI statistic values. The conclusion was same from results at other locations, and for other months, i.e. temporal dependence among drought states cannot be ignored. To account for the dependence in drought states while modeling even one-month droughts, HMMs are needed over the simpler mixture models.

Table 1
Rooting depths (in metres) for crops grown in Indiana over the annual growing season, where symbol '×' represents absence of cultivation (Weaver, 1926; Weaver and Bruner, 1927; Rhoads and Yonts, 1991).

Crop	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Corn	×	×	×	×	0.65	0.9	0.9	0.9	0.9	1	1	×
Soybean	×	×	×	×	0.5	0.76	0.9	1	1.4	1.5	1.8	×
Sorghum	×	×	×	×	×	1.2	1.55	1.65	1.85	1.85	×	×
Alfalfa	×	×	0.13	0.5	0.9	1	1.2	1.5	2	×	×	×
Winter wheat	0.8	1	1.2	1.3	1.3	×	×	×	×	×	0.5	0.6
Double crop-ww*and soybean	0.8	1	1.2	1.3	1.3	0.5	0.9	1	1.5	1.5	0.5	0.6
Double crop-ww* and corn	0.8	1	1.2	1.3	1.3	0.65	0.9	0.9	1	1	0.5	0.6

* winter wheat

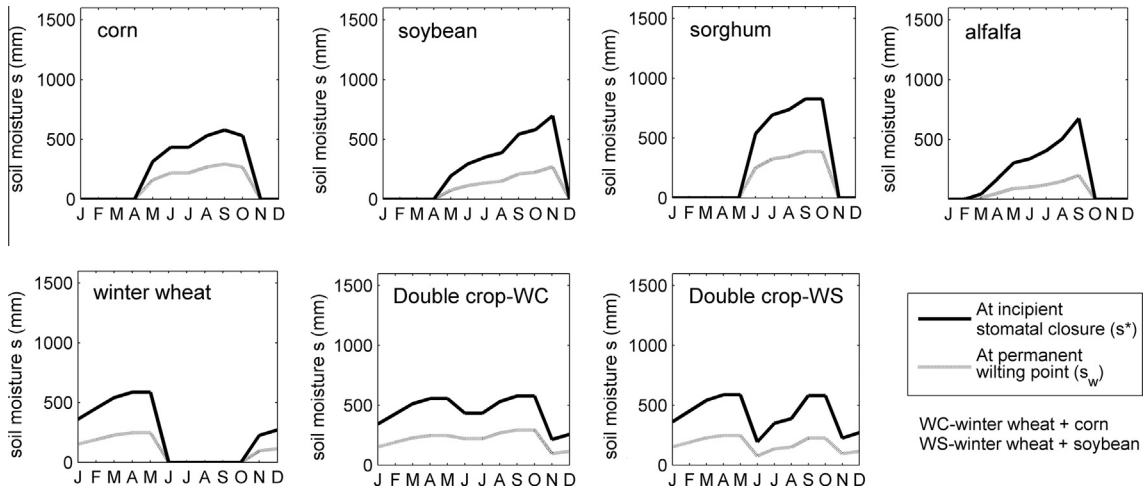

Fig. 4. Monthly soil moisture content values at wilting point (s_w), and at incipient stomatal closure (s^*) for various crops in the study region calculated based on crop growth stage and water requirements.

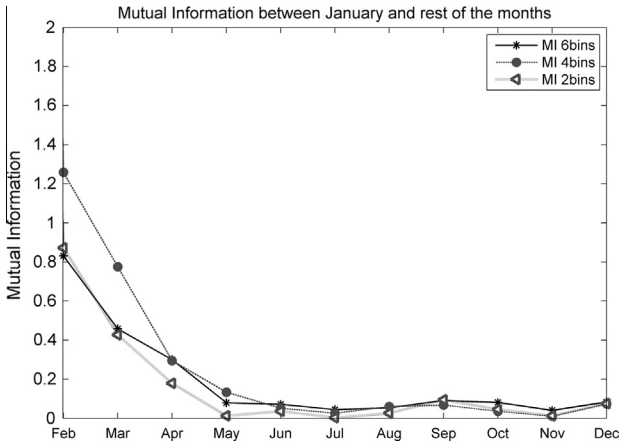

Fig. 5. Mutual information statistic between standardized crop stress function values of January and rest of the months for 2, 4 and 6 bins.

As the number of hidden states increases in HMMs, the corresponding number of model parameters also increases, adding significantly to model complexity and data requirements. In the present study, HMMs with four hidden states are considered for probabilistic assessment of agricultural droughts. These hidden states would represent instances of near normal conditions, moderate, severe and extreme droughts, respectively. Further, as soil moisture changes slowly, the transition probabilities are modeled by a tridiagonal matrix, implying the system could continue in the present state or move to a one-level drier or wetter state over a single time step. These model constraints ensured smaller number of model parameters and more stable results.

4.3. Development of probabilistic drought model

An HMM-based probabilistic drought classification was developed using the crop water stress values at all the 52 grid locations in Indiana as the drought states do share dependence in time. The parameter estimation procedure included initialization and estimation of initial state probabilities, transition probability matrix and beta emission distribution parameters. Scaled HMM (Rabiner, 1989) was used herein to facilitate parameter estimation. The best set of parameters was identified based on maximum log [$P(O|\text{model})$] value from the simulations obtained using random initial values. As noted earlier, a tridiagonal transition matrix was assumed at the second stage of parameter estimation, after having set the order of hidden states in increasing order of drought severity. The best parameter values were then obtained from simulations using random values as transition probabilities, and previously estimated beta emission parameters to initialize the parameters of beta-HMM.

Parameter estimates for the HMM at six locations in Indiana are shown in Table 2 and Fig. 6 as representative samples, for the sake of brevity. These are geographically widely separated points and are denoted by their location identifiers (loc id): 7 (41.25°N, 87.25°W), 9 (41.25°N, 86.25°W), 12 (41.25°N, 84.75°W), 46 (38.25°N, 87.25°W), 35 (39.25°N, 85.75°W), and 44 (38.75°N, 84.75°W) respectively. As expected, in most cases, preference is expressed for continuing in the present state than transitioning to a neighboring state.

The emission distributions at all the six locations in Fig. 6 allow for some statistical interpretation into the drought states. They represent the changing nature of agricultural droughts with spatial locations. At all locations, the emission distributions for near normal conditions have very peaked distributions with a large probability mass concentrated close to $\zeta = 0$. At loc. id 7, as seen in Fig. 6a, the emission distributions for all drought classes have reasonable separation implying that the model is able to resolve these classes with less uncertainty. The peaked probability density functions for near normal and extreme drought states at all locations indicate that these categories are classified with higher probabilities. However, for loc. id 7 and 12, high classification uncertainty exists for severe and extreme droughts (Fig. 6, plots a, c), as the emission distributions have more overlap for severe and extreme drought classes. Consequently, higher transition probabilities exist for transition of

Table 2
Estimated hidden Markov model probabilities- initial state (π_i) and transition state probabilities, and beta emission distribution parameters α and β associated with the four drought states (1-near normal, 2-moderate, 3-severe and 4-extreme) for six locations in Indiana.

		(a) loc id 7				(b) loc id 9				(c) loc id 12			
Drought State →		*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
	π_i	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Transition Probabilities	*1*	0.80	0.20	0.00	0.00	0.80	0.20	0.00	0.00	0.81	0.19	0.00	0.00
	2	0.21	0.62	0.17	0.00	0.34	0.27	0.39	0.00	0.23	0.60	0.18	0.00
	3	0.00	0.54	0.01	0.45	0.00	0.34	0.57	0.09	0.00	0.48	0.01	0.51
	4	0.00	0.00	1.00	0.00	0.00	0.00	0.53	0.47	0.00	0.00	1.00	0.00
	α	1	2	17	35	1	2	9	47	2	5	15	48
	β	221	3	7	11	188	5	7	17	37	9	10	28
		(d) loc id 46				**(e) loc id 35**				**(f) loc id 44**			
	π_i	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Transition Probabilities	*1*	0.81	0.19	0.00	0.00	0.80	0.20	0.00	0.00	0.80	0.20	0.00	0.00
	2	0.25	0.55	0.20	0.00	0.27	0.47	0.26	0.00	0.22	0.60	0.18	0.00
	3	0.00	0.32	0.67	0.01	0.00	0.34	0.59	0.07	0.00	0.33	0.63	0.05
	4	0.00	0.00	0.51	0.49	0.00	0.00	0.57	0.43	0.00	0.00	0.46	0.54
	α	1	4	14	13	1	3	9	34	2	4	19	43
	β	41	9	7	2	130	9	7	12	76	12	15	14

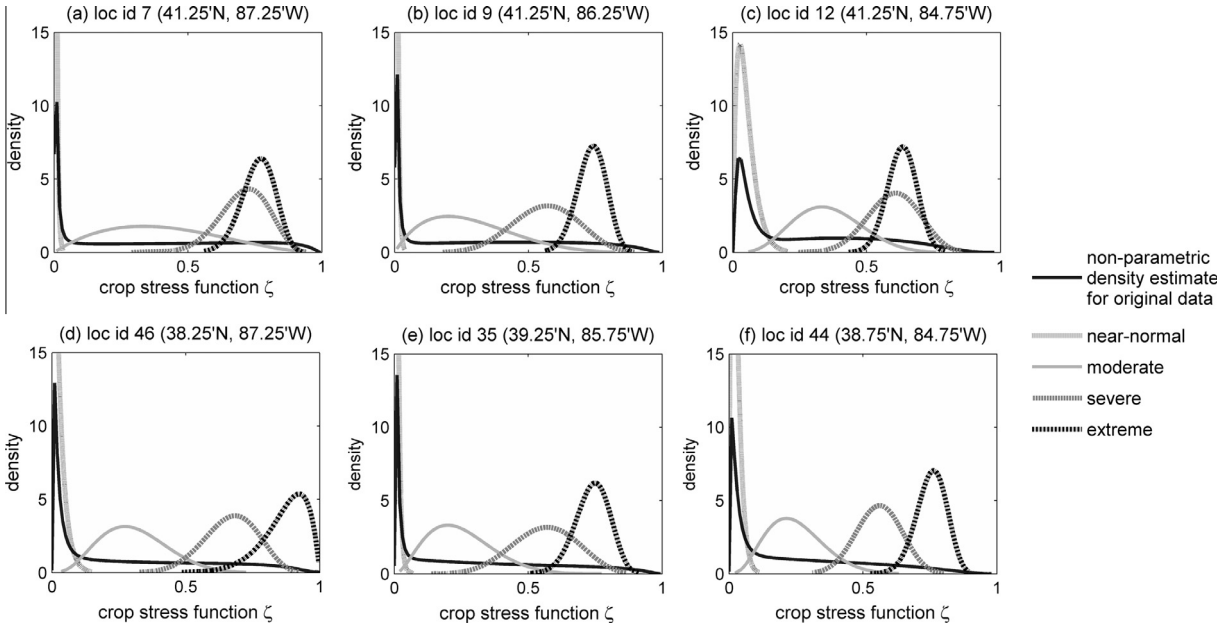

Fig. 6. Estimated emission densities (beta distribution probability density functions) for six locations across Indiana.

extreme drought to severe drought state at these two locations (Table 2 a, c). The moderate drought class, on the other hand has very little overlap in all the six cases, implying less uncertainty in its classification.

Figs. 7a and b show probabilistic classification of drought states provided by the crop water stress-based drought index in HMM framework. Results from only two locations for an example 12 year period 2001–2012 are shown here for the sake of brevity. The height of each bar in the plot represents the probability of a particular drought state in a particular month. While the lighter shade represents a near-normal condition, the darker ones represent increasing severity of drought induced by crop water stress. For instance in Fig. 7a, July 2012 at loc id 7 had the following drought probabilities: 98.2% of being in severe drought, and 1.6% and 0.2% of being in moderate and extreme states, respectively. Similarly, HMM-based classification for August 2012 at loc id 35 indicates 72.5% and 27.5% probabilities of being in extreme and severe states, respectively (Fig. 7b). In contrast to popular indices such as SPI, SPEI and PDSI, the probabilistic drought state classification offered by the proposed index addresses uncertainty in drought characterization. Comparisons with these indices are discussed in the following section.

4.4. Comparison with popular drought indices

Most drought studies have relied on the PDSI (based on a soil water balance equation), and the SPI (based on a precipitation time series). Instead of PDSI, a self-calibrating PDSI (SC-PDSI) that can account for the regional variability in climate (Wells et al., 2004) was used for comparison purposes. As the PDSI is not multiscalar, and a fully meteorological-based SPI cannot provide any indication of crop water stress, both these indices are incapable of evaluating agricultural droughts at different locations in Indiana. SPEI-based analyses conducted by Vicente-Serrano et al. (2012) show that SPEI possesses good correlation with soil moisture in most of the sites in North America. The SPEI computation uses monthly precipitation minus potential evapotranspiration, i.e. a water balance deficit data series, that is aggregated at different time scales as in SPI (McKee et al., 1993), and standardized using a three-parameter

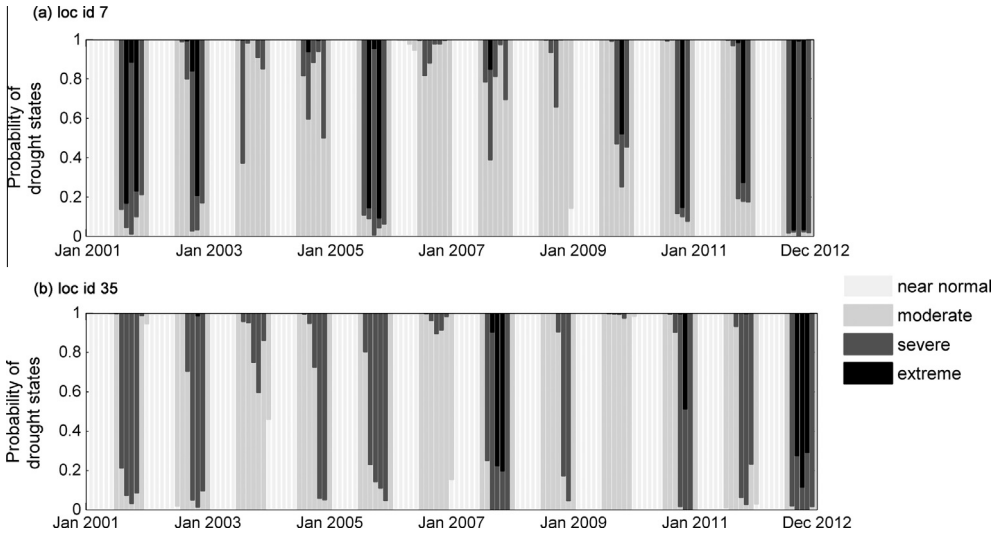

Fig. 7. Probabilistic classification of agricultural droughts during 2001–2012 period at (a) loc id 7 and (b) loc id 35 using the proposed crop stress-based index.

log-logistic distribution (Vicente-Serrano et al., 2010). SPEI time series were computed using SPEI calculator program developed by Beguería and Vicente Serrano (2009); inputs for the program include precipitation and temperature data, as well as the latitude of the selected location. Therefore, for comparison purposes, the SPEI index is also utilized, and relative merits and demerits of all the four indices are evaluated.

Drought category classifications for all indices used in the study are listed in Table 3. Unlike SPI and SC-PDSI, drought categorization with SPEI is fairly recent (Yu et al., 2013). For the proposed HMM-based index, there is no hard classification, and the probability associated with each drought state at a given time can be obtained. For comparison purposes, the predominance of a particular state is indicated when the probability of falling in it exceeds the sum of probabilities of falling in the other states.

Figs. 8 and 9 show the probabilistic monthly drought classification offered by HMM and the corresponding SPEI, SC-PDSI and SPI index values during an example 20 year period – from 1983 to 2003 at loc. id 7 and 35 respectively. The HMM-based method yields probabilities associated with each drought category, thus providing a basis for assessing classification uncertainty, unlike SPI, SPEI or

Table 3
Drought category classification of the drought indices used in this study: proposed crop stress-based index, SPI, SPEI and SC-PDSI.

Hidden state	Drought definition	SPI (McKee et al., 1993)	SPEI (Yu et al., 2013)	SC-PDSI (Wells et al., 2004)
1	Near normal	+1 to −0.99	+1 to −0.99	+0.5 to −0.99
2	Moderate drought	−1 to −1.49	−1 to −1.49	−1 to −2.99
3	Severe drought	−1.5 to −1.99	−1.5 to −1.99	−3 to −3.99
4	Extreme drought	Less than −2	Less than −2	Less than −4

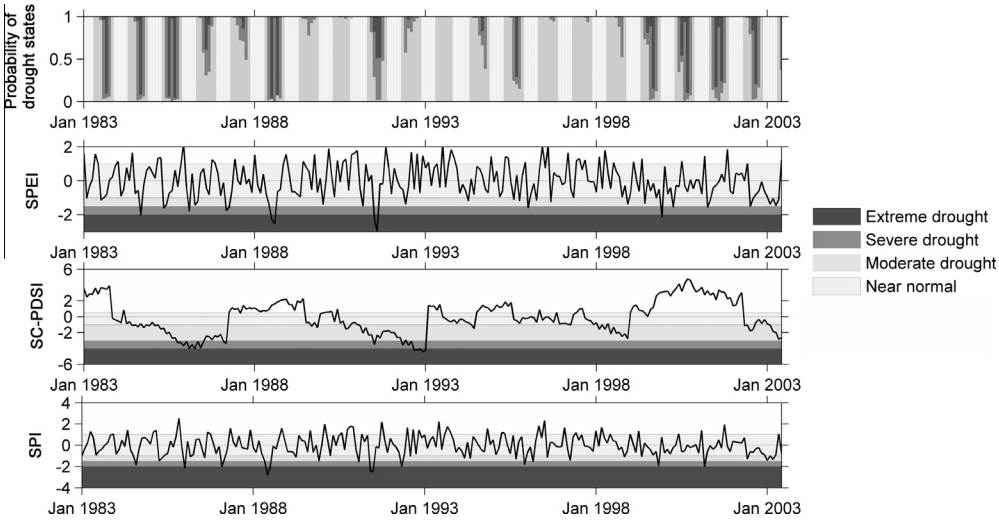

Fig. 8. Comparison between HMM-based agricultural drought index, SPEI, SC-PDSI and SPI values for location id 7 (lat/lon 41.25°, −87.25°) during the 1983–2003 period.

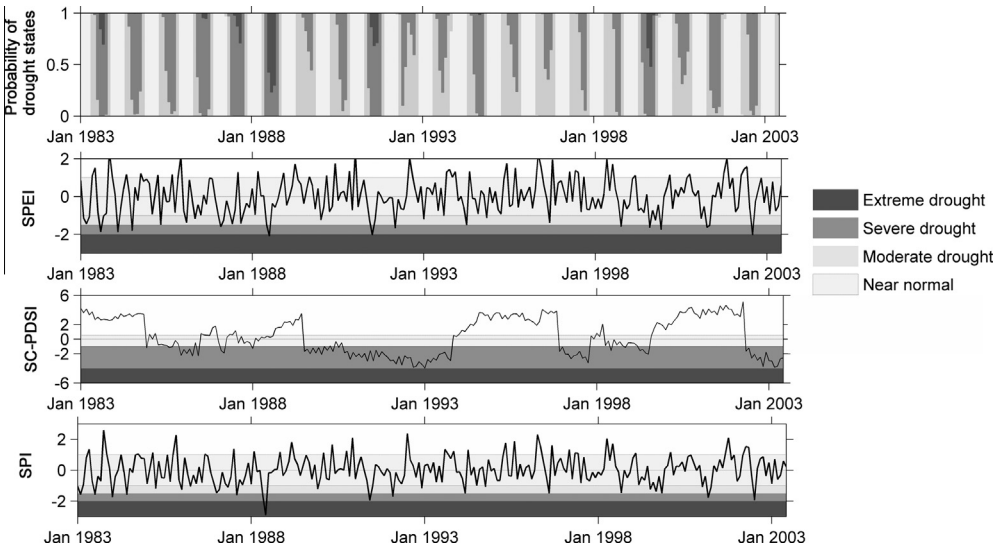

Fig. 9. Comparison between HMM-based agricultural drought index, SPEI, SC-PDSI and SPI values for location id 35 (lat/lon 39.25°, −85.75°) during the 1983–2003 period.

SC-PDSI. At loc. id 7, (Fig. 8), few extreme and severe agricultural drought events are identified in the years 1983–1985, 1988, 1995, 1999–2002, according to the proposed crop stress-based index. SPI and SPEI reported extreme droughts in 1984, 1988, 1991–1992, 1999–2000. On the other hand, SC-PDSI detected very few extreme events during this period, in 1985–1986 and 1993. Severe droughts according to the SPI and SPEI indices, occurred in 1985–1987, 1992 and 2000, and are identified by the proposed index as well. All the indices suggest that near normal to moderate drought conditions are more prevalent in loc id 7. In Fig. 9, at loc. id 35, very few extreme events are suggested in 1988 and 1999 by the proposed index, and severe drought events are more prevalent. SPI and SPEI projected extreme droughts for years 1988, 1991, and 2002, whereas SC-PDSI reported extremes in 1992–1993 and 2003. Moderate drought events are observed frequently during June–September months. The results at these two locations therefore suggest that the developed probabilistic index is capable of identifying agricultural drought events that may not be captured by the SPI, SPEI or SC-PDSI, especially during the months of May–October, the growing season for most of the crops. Additionally, the probabilities assigned to each drought category in the HMM-based probabilistic classification reflect the uncertainty involved in drought identification. The other indices were not designed for this capability.

Since different indices are designed for different purposes and yield different information, the superiority of any one index over others cannot be established. Comparisons between results from different indices may imply robustness if results are consistent. For example, the number of extreme events detected by the proposed index and SC-PDSI during the data period 1948–2012 in Indiana is shown in Fig. 10, pooling drought information from across all the 52 locations in Indiana. According to the proposed crop stress-based index, northern Indiana is relatively more prone to extreme agricultural droughts, while southwest Indiana has had relatively few instances over the data period. The drought maps for extreme events from the proposed index and SC-PDSI are markedly different, suggesting that different indices may lead to different conclusions. There is some agreement in the extreme drought occurrences suggested by the proposed index and SC-PDSI for south-eastern,

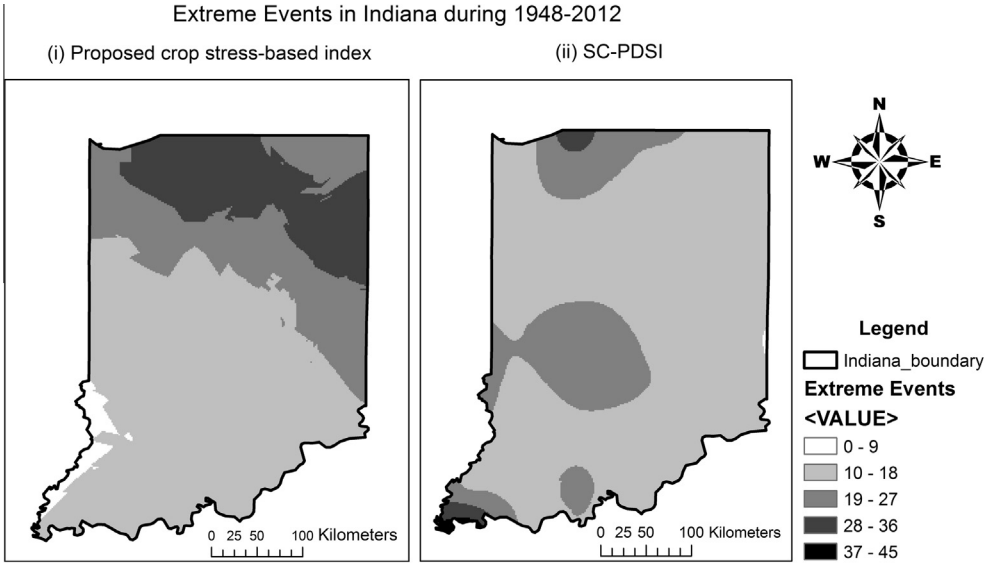

Fig. 10. Extreme drought category maps for Indiana under (i) the proposed crop stress-based index, and (ii) SC-PDSI. Darker shades correspond to increased frequency of extreme droughts during 1948–2012.

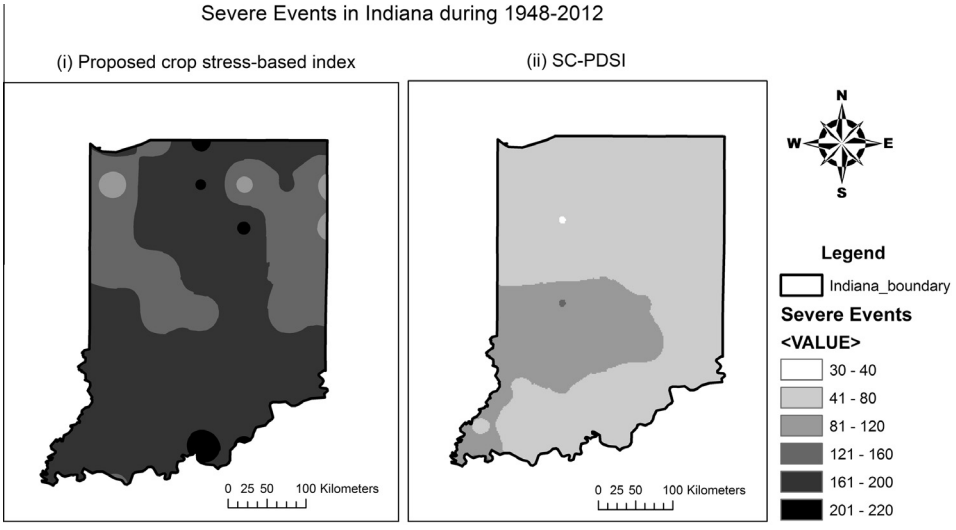

Fig. 11. Severe drought category maps for Indiana under (i) the proposed crop stress-based index, and (ii) SC-PDSI. Darker shades correspond to increased frequency of severe droughts during 1948–2012. SC-PDSI reports a smaller range of occurrences compared to the proposed index.

south-central and central Indiana, but the proposed index would suggest that the state is more prone to extreme droughts.

Similarly, severe drought event maps were constructed for Indiana using the two indices and are shown in Fig. 11. The ranges of number of severe events during the period 1948–2012 identified by the proposed index and SC-PDSI are vastly different. The proposed index reported numerous instances of severe droughts all over the region, far more than those identified by SC-PDSI. The SC-PDSI maps in Figs. 10 and 11 consistently indicate that west and central Indiana have experienced high frequency of extreme and severe category droughts over the 1948–2012 period. However, the proposed index suggests that central and southern Indiana are highly prone to severe droughts. Again, we emphasize that one index is not superior to the other, just that different indices may yield different results implying the choice of an index for drought classification should be based on the specific needs of the user. An evaluation of relative drought-proneness of a region cannot be evaluated by SPI and SPEI as all locations are allocated the same probability of a drought class by definition.

5. Conclusions

A probabilistic agricultural drought index that is based on crop water needs was formulated within a graphical model (HMM) framework, where hidden states represent different drought categories (from near normal to extreme droughts). The monthly soil moisture-based crop water stress function estimated in the study was found to have temporal dependence between drought states, thus suggesting the use of HMMs over simpler mixture models. Crop water stress was modeled using HMMs with a tridiagonal transition matrix and beta emission densities to develop a probabilistic model based on a bounded stress function.

Retrospective comparison of drought events of an example 20 year period (1983–2003) suggested by the proposed model and indices such as SPEI, SC-PDSI and SPI indicated fairly good agreement over agricultural drought conditions. Given that consistent definitions of corresponding SPI, SPEI and SC-PDSI index values for each drought state in the HMM framework—near normal, moderate, severe and extreme droughts are not available, direct comparisons could not be made. Additionally, extreme and severe drought category maps were developed for whole of Indiana using results from the proposed crop water stress-based index, to study the spatial variation of drought-proneness of the study region.

The following observations are made regarding the probabilistic agricultural drought index developed in this paper:

i. Drought severity category is defined differently for each location by the HMM. Drought states evolve based on the historical crop water stress time series at each location, and hence, an averaged or aggregated assessment for a region cannot be considered accurate.
ii. The tridiagonal transition matrix assumption adopted in HMMs in this study holds good for smooth transitioning of drought states and facilitates robust parameter estimation. However, sudden drought transitions that occur in the case of flash droughts may not be well captured by the model under this assumption.
iii. The transition trends and emission distributions are not similar over Indiana. Results tend to be site-specific, suggesting the need for advanced regionalization studies for regional agricultural drought outlook.
iv. For comparisons with existing drought indices, the predominant drought category after probabilistic classification was defined as the one whose probability of occurrence was more than the sum of probabilities of droughts in all other categories. In the event that no drought category is dominant, the classification uncertainty is likely to be high, i.e. multiple drought categories are about equally likely. In the present study, predominant drought categories were distinctly identified over the study area.
v. Comparison of indices indicated that many drought events during dominant crop growing season (May–October) that were not identified by the SPI, SPEI and SC-PDSI, were revealed by the proposed index.
vi. The spatial variation of propensity of extreme and severe category droughts over Indiana during the 1948–2012 period was examined by the proposed crop stress-based index (Figs. 10 and 11). Such maps are useful for planning crop cultivation under rain-fed conditions. Since different indices yield different results, the choice of the index should be based on the desired end result. The utility of these maps need to be further explored in identifying regions where certain crops can be cultivated with minimum chances of crop water stress.

The proposed HMM-based drought index enables classifying agricultural droughts in a probabilistic framework unlike the SPI, SC-PDSI or SPEI. The graphical model-based index highlights the inherent uncertainty in drought analysis, and the framework would be useful in developing reliable forecasting models. The crop water stress-based drought index developed using HMM also suggests the need for alternate drought classification regimes that are driven by the data.

Acknowledgements

Studies were supported in part by the National Science Foundation under Grant AGS 1025430, and by USDA NIFA award number 2011-67019-21122. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the USDA.

Appendix A

A.1. Parameter estimation using EM algorithm

Given that observations in the time series and underlying sequence of states are represented as $O = (o_1, o_2, \ldots, o_T)$ and $q = (q_1, q_2, \ldots, q_T)$, respectively, the log likelihood function to be maximized becomes:

$$Q(\mathrm{M}, \mathrm{M}') = \sum_q P(O, q|\mathrm{M}') \log P(O, q|\mathrm{M}) \tag{A1}$$

where M represents the new set of model parameters and M′ the previous/initial set of values. If we define the probability $P(O, q|\mathrm{M})$ as follows:

$$P(O, q|\mathrm{M}) = \pi_{q_1} \prod_{t=2}^{T} a_{q_{t-1}q_t} b_{q_t}(o_t) \tag{A2}$$

where π, a and b denote initial state, transition and emission probabilities respectively.

Then, Q may be written as:

$$\begin{aligned} Q(\mathrm{M},\mathrm{M}') = &\left[\sum_q P(O,q|\mathrm{M}') \log \pi_{q_1} + \sum_q \sum_{t=2}^{T} \log a_{q_{t-1}q_t} P(O,q|\mathrm{M}') \right. \\ &\left. + \sum_q \sum_{t=1}^{T} \log b_{q_t}(o_t) P(O,q|\mathrm{M}') \right] \\ = &\, Q(I) + Q(II) + Q(III) \end{aligned} \tag{A3}$$

A.2. Estimation of initial state probabilities π_i

The three parts of Eq. (A3) can be used to maximize the Q function; each term is optimized independently to obtain the new set of parameters.

For the first term $Q(I)$ in Eq. (A3), its maximization subject to constraint $\sum_i \pi_i = 1$ to obtain estimation formula for π_i follows:

$$\frac{\partial}{\partial \pi_i}\left(\sum_{i=1}^{K}\log \pi_i\, p(O, q_1 = i|\mathrm{M}') + \lambda_1\left(\sum_{i=1}^{K}\pi_i - 1\right)\right) = 0$$
$$\Rightarrow \pi_i = \frac{P(O,q_1=i|\mathrm{M}')}{P(O|\mathrm{M}')} = \frac{\alpha_1^{*\prime}(i)\beta_1^{*\prime}(i)}{\sum_{i=1}^{K}\alpha_1^{*\prime}(i)\beta_1^{*\prime}(i)} \tag{A4}$$

where λ_1 is the Lagrange multiplier, and functions α^*, β^* are as defined in the forward-backward algorithm of Rabiner (1989). Note that these functions are different from the beta emission parameters.

A.3. Estimation of transition state probabilities a_{ij}

Similar to the previous exercise, maximization of $Q(II)$ subject to constraint $\sum_{j=1}^{K} a_{ij} = 1$ is performed as follows:

$$\frac{\partial}{\partial a_{ij}}\left(\sum_{q}\sum_{t=2}^{T}\log a_{q_{t-1}q_t} P(O,q|\mathrm{M}')\right) + \lambda_2\left(\sum_{j=1}^{K}a_{ij} - 1 = 0\right)$$
$$= \frac{\partial}{\partial a_{ij}}\left(\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=2}^{T}\log a_{ij} P(O, q_{t-1=i}, q_{t=j}|\mathrm{M}')\right) + \lambda_2\left(\sum_{j=1}^{K}a_{ij} - 1 = 0\right)$$
$$\Rightarrow a_{ij} = \frac{\sum_{t=2}^{T}P(O,q_{t-1=i},q_{t=j}|\mathrm{M}')}{\sum_{t=2}^{T}P(O,q_{t-1=i}|\mathrm{M}')} = \frac{\sum_{t=2}^{T}\alpha_{t-1}^{*\prime}(i)a'_{ij}\beta_t^{*\prime}(j)b'_j(o_t)}{\sum_{j=1}^{K}\sum_{t=2}^{T}\alpha_{t-1}^{*\prime}(i)a'_{ij}\beta_t^{*\prime}(j)b'_j(o_t)} \tag{A5}$$

A.4. Estimation of beta emission distribution parameters

Maximizing $Q(III)$ does not involve Lagrange multipliers as there are no constraints for the beta emission distribution parameters α_j, β_j.

$$\frac{\partial}{\partial \alpha_i}\left(\sum_{q}\sum_{t=1}^{T}\log b_{q_t}(o_t) P(O,q|\mathrm{M}')\right) = \left(\sum_{q}\sum_{t=1}^{T}P(O,q|\mathrm{M}')\frac{\partial(\log b_{q_t}(o_t))}{\partial \alpha_j}\right) = 0$$
$$\frac{\partial}{\partial \beta_i}\left(\sum_{q}\sum_{t=1}^{T}\log b_{q_t}(o_t) P(O,q|\mathrm{M}')\right) = \left(\sum_{q}\sum_{t=1}^{T}P(O,q|\mathrm{M}')\frac{\partial(\log b_{q_t}(o_t))}{\partial \beta_j}\right) = 0 \tag{A6}$$

Emission density for beta distribution is $b_j(O_t) = betapdf(o_t, \alpha_j, \beta_j) = \frac{o_t^{\alpha_j-1}(1-o_t)^{\beta_j-1}}{\mathrm{B}(\alpha_j,\beta_j)}$.

Derivatives in Eq. (A6) can be expanded as:

$$\frac{\partial(\log b_{q_t}(o_t))}{\partial \alpha_j} = \frac{1}{b_{q_t}(o_t)}\frac{\partial(b_{q_t}(o_t))}{\partial \alpha_j} = \frac{1}{b_{q_t}(o_t)}\cdot\frac{\partial}{\partial \alpha_j}\left(\frac{o_t^{\alpha_j-1}(1-o_t)^{\beta_j-1}}{\mathrm{B}(\alpha_j,\beta_j)}\right)$$
$$= \frac{(1-o_t)^{\beta_j-1}}{b_{q_t}(o_t)}\left[\frac{\partial o_t^{\alpha_j-1}}{\partial \alpha_j}\cdot\frac{1}{\mathrm{B}(\alpha_j,\beta_j)} + o_t^{\alpha_j-1}\cdot\frac{\partial}{\partial \alpha_j}\left(\frac{1}{\mathrm{B}(\alpha_j,\beta_j)}\right)\right]$$
$$= \frac{(1-o_t)^{\beta_j-1}}{b_{q_t}(o_t)}\left[\frac{o_t^{\alpha_j-1}\log o_t}{\mathrm{B}(\alpha_j,\beta_j)} + o_t^{\alpha_j-1}(-1)\left(\frac{1}{\mathrm{B}(\alpha_j,\beta_j)}\right)^2(\psi(\alpha_j)-\psi(\alpha_j+\beta_j))\mathrm{B}(\alpha_j,\beta_j)\right]$$
$$= \frac{o_t^{\alpha_j-1}(1-o_t)^{\beta_j-1}}{\mathrm{B}(\alpha_j,\beta_j)b_{q_t}(o_t)}\left[\log o_t - \psi(\alpha_j) + \psi(\alpha_j+\beta_j)\right] = \frac{b_{q_t}(o_t)}{b_{q_t}(o_t)}\left[\log o_t - \psi(\alpha_j) + \psi(\alpha_j+\beta_j)\right]$$
$$= \log o_t - \psi(\alpha_j) + \psi(\alpha_j+\beta_j) \tag{A7}$$

$$\frac{\partial(\log b_{q_t}(o_t))}{\partial \beta_j} = \frac{1}{b_{q_t}(o_t)}\frac{\partial(b_{q_t}(o_t))}{\partial \beta_j} = \frac{1}{b_{q_t}(o_t)}\cdot\frac{\partial}{\partial \beta_j}\left(\frac{o_t^{\alpha_j-1}(1-o_t)^{\beta_j-1}}{\mathrm{B}(\alpha_j,\beta_j)}\right)$$
$$= \frac{o_t^{\alpha_j-1}}{b_{q_t}(o_t)}\left[\frac{\partial(1-o_t)^{\beta_j-1}}{\partial \beta_j}\cdot\frac{1}{\mathrm{B}(\alpha_j,\beta_j)} + (1-o_t)^{\beta_j-1}\cdot\frac{\partial}{\partial \beta_j}\left(\frac{1}{\mathrm{B}(\alpha_j,\beta_j)}\right)\right]$$
$$= \frac{o_t^{\alpha_j-1}}{b_{q_t}(o_t)}\left[\frac{(1-o_t)^{\beta_j-1}\log(1-o_t)}{\mathrm{B}(\alpha_j,\beta_j)} + (1-o_t)^{\beta_j-1}(-1)\left(\frac{1}{\mathrm{B}(\alpha_j,\beta_j)}\right)^2(\psi(\beta_j)-\psi(\alpha_j+\beta_j))\mathrm{B}(\alpha_j,\beta_j)\right]$$
$$= \frac{o_t^{\alpha_j-1}(1-o_t)^{\beta_j-1}}{\mathrm{B}(\alpha_j,\beta_j)b_{q_t}(o_t)}\left[\log(1-o_t) - \psi(\beta_j) + \psi(\alpha_j+\beta_j)\right] = \frac{b_{q_t}(o_t)}{b_{q_t}(o_t)}\left[\log(1-o_t) - \psi(\beta_j) + \psi(\alpha_j+\beta_j)\right]$$
$$= \log(1-o_t) - \psi(\beta_j) + \psi(\alpha_j+\beta_j) \tag{A8}$$

Note that $\psi(\cdot)$ denotes digamma function formed during differentiation of beta function $\mathrm{B}(\alpha_j, \beta_j) = \Gamma(\alpha_j)\Gamma(\beta_j)/\Gamma(\alpha_j + \beta_j)$. The derivatives of beta function are determined as follows:

$$\frac{\partial \mathrm{B}(\alpha_j,\beta_j)}{\partial \alpha_j} = \mathrm{B}(\alpha_j,\beta_j)\left[\frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} - \frac{\Gamma'(\alpha_j+\beta_j}{\Gamma(\alpha_j+\beta_j)}\right] = \mathrm{B}(\alpha_j,\beta_j)\left[\psi(\alpha_j) - \psi(\alpha_j+\beta_j)\right]$$
$$\text{Also, } \frac{\partial \mathrm{B}(\alpha_j,\beta_j)}{\partial \beta_j} = \mathrm{B}(\alpha_j,\beta_j)\left[\psi(\beta_j) - \psi(\alpha_j+\beta_j)\right] \tag{A9}$$

Therefore, the emission density parameter estimation problem reduces to solution of following two equations:

$$\sum_{j=1}^{K}\sum_{t=1}^{T}P(O,q|\mathrm{M}')\left[\log(o_t) - \psi(\alpha_j) + \psi(\alpha_j+\beta_j)\right] = 0$$
$$\sum_{j=1}^{K}\sum_{t=1}^{T}P(O,q|\mathrm{M}')\left[\log(1-o_t) - \psi(\beta_j) + \psi(\alpha_j+\beta_j)\right] = 0 \tag{A10}$$

The parameter estimation procedure outlined above was repeated, the log-likelihood increased with every iteration, until the solutions for different unknowns converged. The forward-backward algorithm computations for large datasets involved summation of a large number of terms that exceeded the precision range of computing machines. However, these steps are inevitable for estimation of parameters in HMM. In order to cope with this issue, scaling was performed (Rabiner, 1989).

The posterior probability of being in a particular drought state at time t, that forms the basis for estimating the uncertainty in drought state classification, is given by:

$$P(q_t = i|O, M') = \frac{\alpha_t^{*\prime}(i)\beta_t^{*\prime}(i)}{\sum_{i=1}^{K}\alpha_t^{*\prime}(i)\beta_t^{*\prime}(i)} \tag{A11}$$

References

Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Ann. Math. Stat. 41, 164–171.

Beguería, S., Vicente Serrano, S.M., 2009. SPEI Calculator, http://digital.csic.es/handle/10261/10002.

Bergman, K.H., Sabol, P., Miskus, D., 1988. Experimental indices for monitoring global drought conditions. In: Proceedings of the Thirteenth Annual Climate Diagnostics Workshop. U.S. Dept. of Commerce, Cambridge, MA, pp. 190–197.

Bolten, J.D., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A., 2010. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. Sel. Top. Appl. Earth Observations Remote Sens., IEEE J. 3, 57–66.

Burget, L., Schwarz, P., Agarwal, M., Akyazi, P., Feng, K., Ghoshal, A., Glembek, O., Goel, N., Karafiat, M., Povey, D., Rastrow, A., Rose, R.C., Thomas, S., 2010. Multilingual acoustic modeling for speech recognition based on subspace Gaussian Mixture Models. In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 4334–4337. http://dx.doi.org/10.1109/ICASSP.2010.5495646.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.

Crouse, M.S., Nowak, R.D., Baraniuk, R.G., 1998. Wavelet-based statistical signal processing using hidden Markov models. Signal Process., IEEE Trans. 46, 886–902.

Dai, A., Trenberth, K.E., Qian, T., 2004. A global dataset of palmer drought severity index for 1870–2002: relationship with soil moisture and effects of surface warming. J. Hydrometeor 5, 1117–1130.

Denmead, O.T., Shaw, R.H., 1960. The effects of soil moisture stress at different stages of growth on the development and yield of corn. Agron. J. 52, 272–274.

Evans, R., Cassel, D.K., Sneed, R.E., 1996. Soil, water, and crop characteristics important to irrigation scheduling. AG 452-1, North Carolina Cooperative Extension Service, NC.
Fan, Y., van den Dool, H., 2004. Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. J. Geophys. Res. 109, D10102.
Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84, 111–123.
Hocaoğlu, F.O., Gerek, Ö.N., Kurban, M., 2010. A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models. J. Wind Eng. Ind. Aerodyn. 98 (8–9), 472–481.
Holt, R.F., Timmons, D.R., Voorhees, W.B., Van Doren, C.A., 1964. Importance of stored soil moisture to the growth of corn in the dry to moist subhumid climatic zone. Agron. J. 56, 82–85.
Hsiao, T.C., 1973. Plant responses to water stress. Annu. Rev. Plant Physiol. 24, 519–570.
Huang, J., van den Dool, H.M., Georgarakos, K.P., 1996. Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts. J. Clim. 9, 1350–1362.
Ihler, A.T., Kirshner, S., Ghil, M., Robertson, A.W., Smyth, P., 2007. Graphical models for statistical inference and data assimilation. Physica D 230, 72–87.
Jordan, M.I., 2004. Graphical models. Stat. Sci. 19, 140–155.
Karamouz, M., Torabi, S., Araghinejad, S., 2004. Analysis of hydrologic and agricultural droughts in Central Part of Iran. J. Hydrol. Eng. 9, 402–414.
Kerr, E., 2012. Brutal drought depresses agriculture, thwarting US and Texas economies. Southwest Econ. Fourth Quarter, 10–13.
Keyantash, J., Dracup, J.A., 2002. The quantification of drought: an evaluation of drought indices. Bull. Am. Meteor. Soc. 83, 1167–1180.
Koster, R.D., Suarez, M.J., 2001. Soil moisture memory in climate models. J. Hydrometeor 2, 558–570.
Laio, F., Porporato, A., Fernandez-Illescas, C.P., Rodriguez-Iturbe, I., 2001. Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: IV. Discussion of real cases. Adv. Water Resour. 24, 745–762.
Lakshmi, V., Piechota, T., Narayan, U., Tang, C., 2004. Soil moisture as an indicator of weather extremes. Geophys. Res. Lett. 31. http://dx.doi.org/10.1029/2004GL019930.
Leggetter, C.J., Woodland, P.C., 1995. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Comput. Speech Lang. 9, 171–185.
Leu, S.-S., Adi, T.J.W., 2011. Probabilistic prediction of tunnel geology using a Hybrid Neural-HMM. Eng. Appl. Artif. Intell. 24, 658–665.
Liu, W.T., Kogan, F.N., 1996. Monitoring regional drought using the Vegetation Condition Index. Int. J. Remote Sens. 17, 2761–2782.
Maity, R., Sharma, A., Nagesh Kumar, D., Chanda, K., 2013. Characterizing drought using the reliability-resilience-vulnerability concept. J. Hydrol. Eng. 18, 859–869.
Madadgar, S., Moradkhani, H., 2014. Spatio-temporal drought forecasting within Bayesian networks. J. Hydrol. 512, 134–146.
Madadgar, S., Moradkhani, H., 2013. A bayesian framework for probabilistic seasonal drought forecasting. J. Hydrometeorol. 14, 1685–1705.
Mallya, G., Tripathi, S., Kirshner, S., Govindaraju, R.S., 2013a. Probabilistic assessment of drought characteristics using hidden markov model. J. Hydrol. Eng. 18, 834–845.
Mallya, G., Zhao, L., Song, X.C., Niyogi, D., Govindaraju, R.S., 2013b. 2012 Midwest drought in the United States. J. Hydrol. Eng. 18, 737–745.
Manabe, S., Delworth, T., 1990. The temporal variability of soil wetness and its impact on climate. Clim. Change 16, 185–192.
Margulis, S.A., McLaughlin, D., Entekhabi, D., Dunne, S., 2002. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. Water Resour. Res. 38 (12), 35–1–35–18.
McKee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology, pp. 179–183.
Meyer, S.J., Hubbard, K.G., Wilhite, D.A., 1993. A Crop-specific drought index for Corn: II. application in drought monitoring and assessment. Agron. J. 85, 396–399.
Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216.
Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. Agric. For. Meteorol. 133, 69–88.
NDMC-UNL, 2012. The U.S. Drought Monitor is jointly produced by the National Drought Mitigation Center at the University of Nebraska-Lincoln, the United States Department of Agriculture, and the National Oceanic and Atmospheric Administration. <http://droughtmonitor.unl.edu/MapsandDataServices/MapService.asp>.
Oglesby, R.J., Erickson, D.J., 1989. Soil moisture and the persistence of North American drought. J. Climate 2, 1362–1380.
Palmer, W.C., 1965. Meteorological drought. Weather Bureau Research Paper No. 45. U.S. Dept. of Commerce, Washington, DC, 58pp.
Palmer, W.C., 1968. Keeping track of crop moisture conditions, nationwide: the new crop moisture index. Weatherwise 21, 156–161.
Porporato, A., Laio, F., Ridolfi, L., Rodriguez-Iturbe, I., 2001. Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: III. Vegetation water stress. Adv. Water Resour. 24, 725–744.
Potop, V., Možný, M., Soukup, J., 2012. Drought evolution at various time scales in the lowland regions and their impact on vegetable crops in the Czech Republic. Agric. For. Meteorol. 156, 121–133.
Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77, 257–286.
Rhoads, F.M., Yonts, C.D., 1991. Irrigation scheduling for corn—Why and how. In: National Corn Handbook, NCH-20. USDA, Washington, DC.
Rodriguez-Iturbe, I., D'Odorico, P., Porporato, A., Ridolfi, L., 1999a. On the spatial and temporal links between vegetation, climate, and soil moisture. Water Resour. Res. 35, 3709–3722.
Rodríguez-Iturbe, I., D'Odorico, P., Porporato, A., Ridolfi, L., 1999b. Tree-grass coexistence in Savannas: the role of spatial dynamics and climate fluctuations. Geophys. Res. Lett. 26, 247–250.
Scholes, R.J., Walker, B.H., 1993. An African Savanna: Synthesis of the Nylsvley Study. Cambridge Univ. Press, Cambridge, UK.
Sheffield, J., Goteti, G., Wen, F., Wood, E.F., 2004. A simulated soil moisture based drought analysis for the United States. J. Geophys. Res.: Atmos. 109. http://dx.doi.org/10.1029/2004JD005182.
Sheffield, J., Wood, E.F., 2008. Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. J. Clim. 21, 432–458.
Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., 2013. Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. Sci. Total Environ. 443, 93–103.
Tang, C., Piechota, T.C., 2009. Spatial and temporal soil moisture and drought variability in the Upper Colorado River Basin. J. Hydrol. 379, 122–135.
Tolk, J.A., 2003. Soils, permanent wilting points. Encyclopedia Water Sci.. http://dx.doi.org/10.1081/E-EWS 120010337, Marcel Dekker Inc.
Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., Hopmans, J.W., 2008. On the value of soil moisture measurements in vadose zone hydrology: a review. Water Resour. Res. 44, W00D06. http://dx.doi.org/10.1029/2008WR006829.
Vicente-Serrano, S.M., Beguería, S., Lorenzo-Lacruz, J., Camarero, J.J., López-Moreno, J.I.Azorin-Molina, C., Revuelto, J., Morán-Tejeda, E., Sanchez-Lorenzo, A., 2012. Performance of drought indices for ecological, agricultural, and hydrological applications. Earth Interact. 16, 1–27. http://dx.doi.org/10.1175/2012EI000434.1.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718.
Weaver, J.E., Bruner, W.E., 1927. Root Development of Vegetable Crops, first ed. McGraw-Hill Book Company, Inc., New York. <http://www.soilandhealth.org/01aglibrary/010137veg.roots/010137toc.html>.
Weaver, J.E., 1926. Root Development of Field Crops. McGraw-Hill Book Company, Inc., New York. <http://www.soilandhealth.org/01aglibrary/010139field croproots/010139toc.html>.
Wells, N., Goddard, S., Hayes, M.J., 2004. A self-calibrating Palmer drought severity index. J. Clim. 17, 2335–2351.
Wu, J., He, B., Lü, A., Zhou, L., Liu, M., Zhao, L., 2011. Quantitative assessment and spatial characteristics analysis of agricultural drought vulnerability in China. Nat. Hazards 56, 785–801.
Yau, C., Papaspiliopoulos, O., Roberts, G.O., Holmes, C., 2011. Bayesian non-parametric hidden Markov models with applications in genomics. J. Royal Stat. Soc.: Ser. B (Statistical Methodology) 73 (1), 37–57.
Yu, H., Choo, Z., Uy, W.I.T., Dauwels, J., Jonathan, P., 2012. Modeling extreme events in spatial domain by copula graphical models. In: 2012 15th International Conference On Information Fusion (FUSION), pp. 1761–1768.
Yu, M., Li, Q., Hayes, M.J., Svoboda, M.D., Heim, R.R., 2013. Are droughts becoming more frequent or severe in China based on the Standardized Precipitation Evapotranspiration Index: 1951–2010? Int. J. Climatol.. http://dx.doi.org/10.1002/joc.3701.
Zhang, H., Zhang, W., Palazoglu, A., Sun, W., 2012. Prediction of ozone levels using a Hidden Markov Model (HMM) with Gamma distribution. Atmos. Environ. 62, 64–73.